基于节点属性的社区发现博弈算法 *

张贤坤，任 静，刘渊博，苏 静

(天津科技大学 计算机科学与信息工程学院，天津 300457)

摘 要：近年来，高质量社区的挖掘和发现已经成为社会网络研究的一个热点。提出一种基于节点属性的社区发现博弈算法 G_NA（game algorithm based on node attributes for community detection）。将社区发现的过程看做网络中节点的博弈，当所有节点都不能提高自身收益时，博弈结束。首先，G_NA 提出基于节点度属性的收益函数；然后，在迭代过程中，节点按照重要度从大到小排序，并依次选择策略提高收益；最后，将提出的算法与现有算法分别在不同的真实网络和人工网络上进行对比实验，结果表明提出的算法优于其他算法。

关键词：社区发现；博弈；节点属性；收益函数

中图分类号：TP393 doi: 10.19734/j.issn.1001-3695.2018.05.0444

Game algorithm based on node attributes for community detection

Zhang Xiankun, Ren Jing, Liu Yuanbo, Su Jing

(*School of Computer Science & Information Engineering*, *Tianjin University of Science & Technology*, *Tianjin* 300457, *China*)

Abstract: In recent years, the detection of high-quality communities has become a hot topic in social network research. This paper proposes a game algorithm based on node attributes for community detection (G_NA) . This paper regards the process of community detection as the game of nodes in the network. When all nodes cannot improve their own utilities, the game ends. First, G_NA proposes a utility function based on the node degree attribute. Then, during the iteration, the nodes update the strategies in a sequence of node influence from large to small to increase their utilities. Finally, the proposed algorithm is compared with the existing algorithm in different real networks and artificial networks. The results show that the proposed algorithm is superior to other algorithms.

Key words: community detection; game; node attributes; utility function

0 引言

复杂网络是复杂系统的抽象，现实中许多复杂系统都可以用复杂网络的相关特性进行描述和分析。用图中的节点表示系统中的个体，边表示个体之间的关系[1~3]。其中社区结构是复杂网络中的一个普遍特征，整个网络由许多个社区组成。同一社区内的节点与节点之间的连接很紧密，而社区与社区之间的连接比较稀疏。若任意两个社区的节点集合的交集均为空，则是非重叠社区；否则为重叠社区。对复杂网络的研究一直是很多领域的研究热点，如蛋白质结构分析、城市道路建设以及社会营销领域。随着微博、Facebook 等社交软件的出现，对于复杂网络社区结构的研究更重要。

目前社区发现的算法主要有谱二分法、K-L 算法等图分割算法；GN 算法等分裂方法；模拟退火、极值优化等基于模块度的方法等。其中 K-L 算法必须预先制定两个社区的大小，否则不会得到正确的划分结果，因此实际应用范围很小。谱二分法一次只能划分两个社区，如果需要划分多个社区，则需要多次迭代。所以效率不高，算法准确度也会降低。GN 算法是一个删除边的算法，本质是基于聚类中的分裂思想，在原理上是使用边介数作为相似度的度量方法。在 GN 算法中，每次都会选择边介数高的边删除，进而网络分裂速度远快于随机删除边时的网络分裂。GN 算法不知道最后会有多少个社区；在计算边介数的时候可能会有很对重复计算最短路径的情况。Newman 提出模块度 Q 后，研究者提出模拟退火、极值优化等基于模块度的方法，将社区划分问题转化为优化问题，进而寻找一个目标函数的最优解。

2010 年，Chen 等人[4]首次将博弈论用于社区发现，提出基于博弈论的社区发现算法。但是算法没有考虑节点自身的属性，随机选择初始节点造成实验结果随机性很大。本文在 Chen 等人研究的基础上，提出基于节点属性的社区发现博弈算法 G_NA。首先，本文提出基于节点度属性的收益函数，该收益函数是对 Chen 等收益函数的一种修正；然后，在迭代过程中，节点按照重要度从大到小排序，并依次选择策略提高收益；最后，将本文提出的算法与现有算法分别在不同的真实网络和人

收稿日期：2018-05-09；**修回日期：**2018-06-29 **基金项目：**国家自然基金资助项目（61702367）；天津市教委科研计划项目（2017KJ033）

作者简介：张贤坤，男，教授，博士，主要研究方向为智能信息处理、社会网络分析；任静，女，硕士研究生，主要研究方向为智能信息处理（renjing@mail.tust.edu.cn）；刘渊博，女，硕士研究生，主要研究方向为智能信息处理；苏静，女，副教授，主要研究方向为智能信息处理.

工网络上进行对比实验，结果表明本文的算法优于其他算法。本文的主要贡献有：a）提出了一种基于节点属性的社区发现博弈算法；b）提出了一个基于节点度属性的收益函数；c）博弈中将节点按照重要度从大到小排序进行策略选择；d）本文算法对要检测的社区的大小和数量没有限制。

1 相关工作

1.1 博弈论简介

1.1.1 博弈论的定义

博弈论[5]是关于参与人策略相互作用的理论。博弈论考虑博弈中个体的预测行为及实际行为，并研究它们的优化策略。当个体利益存在冲突时，每个个体所获得的利益除了取决于自己所获取的行动，还依赖于其他个体采取的行动，每个个体都需要针对对方的行为选择做出对自己最有利的反应。

1.1.2 博弈论的要素

博弈论要素包括参与人、策略、收益、均衡等，其中参与人、策略及收益是最基本的博弈要素。参与人是博弈中的决策主体，其通过选择策略最大化自己的收益，参与人可以是自然人、团体、自然等。

参与人 i 的策略集 $S_i=\{s_i\}$ 是其可行策略的集合。策略组合 $s=(s_1,s_2,...s_n)$ 是由参与博弈的每个参与人选择一个策略所组成的一个有序集。

1.1.3 纳什均衡

纳什均衡[6]是指这样的策略组合：为了最大化自己的收益，每个参与人所采取的策略一定是关于其他参与人所采取策略的最佳策略。因此作为理性参与人，没有一个参与人会轻率地偏离这个策略组合而使自己利益受损。

定义 1 纳什均衡。在有 n 个参与人的博弈 $G=\{S_1,S_2...S_n;u_1,u_2,...u_n\}$ 中，策略组合 $s^*=(s_1^*,s_2^*,...s_n^*)$ 是一个纳什均衡，如果对于每一个 i，s_i^* 是给定其他参与人的选择，$S-i^*=(s_1^*,...s_{i-1}^*,s_{i+1}^*...s_n^*)$ 的情况下，第 i 个人的最优策略，即 $u_i(s_i^*,s_{-i}^*)\geq u_i(s_i,s_{-i}^*)$，对所有的 $i\in\Gamma$ 或者用另一种表示方式，s_i^* 是 $s_i^*\in\arg u_i(s_1^*,...s_{i-1}^*,s_i,s_{i+1}^*...s_n^*)$，$i=1,2,...,n$ 的最大化问题的解。

因此，如果一个策略组合中，不存在一个参与人可以单方面改变自己的策略而提高收益，这个策略组合就是纳什均衡。

1.1.4 潜在博弈与潜在函数

潜在博弈[7]是一种特殊类型的博弈方法，收敛性是它的一个重要性质。对于潜在博弈，存在一个局部线性的潜在函数 P，能反映参与者 i 策略单独发生改变时收益函数的变化量。形式化描述为

$$u_i\left(s_i,s_{-i}\right)-u_i\left(s_i^{'},s_{-i}\right)\geq\omega_i\left(P\left(s_i,s_{-i}\right)-P\left(s_i^{'},s_{-i}\right)\right) \tag{1}$$

其中：$s_i,s_i^{'}\in S_i$；$P:S\to R$；ω_i 表示参与者 i 的权值。

纳什均衡规定：潜在博弈至少存在一个纳什均衡点，且满足 $a^*=\arg\max P(a)$ 的 a^* 都是纳什均衡点。

1.2 Game 算法

2010 年，Chen 等人提出一种基于博弈论的社区发现算法，将博弈论用于社区检测。其算法的主要思想是：将复杂网络的社区形成过程建模为社区形成博弈。将网络中的每个个体视为自私的参与者。每个节点通过加入、离开和转换社区提高自己的收益，直到算法达到纳什均衡，此时得到的网络结构为最终的社区结构。这个框架反映了现实网络中社区的有机形成过程。算法的收益函数是由增益函数和损失函数构成。

Chen 等算法中的增益函数是

$$Q_i(L)=\frac{1}{2m}\sum_{j\in[n]}(A_{ij}\hat{\delta}(i,j)-\frac{d_id_j}{2m}\cdot|L_i\cap L_j|) \tag{2}$$

Chen 等算法中的损失函数是

$$g_i(L)=(|L_i|-1)\cdot c \tag{3}$$

Chen 等算法的基本流程是：

a）初始化每个节点为单独社区；

b）随机选择节点选择其最佳策略；

c）循环步骤 b），直到达到纳什均衡。

Chen 等算法的不足之处有：

a）节点选择策略顺序随机，造成实验划分结果不稳定；

b）没有考虑节点自身属性对节点在网络中选择策略时的权利及参与性。

在 Chen 等人算法的基础上，很多研究者做了算法的改进，主要是收益函数和策略方向的改进。例如，2011 年，Alvari 等人[8]提出基于节点相似度的收益函数，并提出一种新的节点相似度，但是实验效果并不好。2017 年，Zhou 等人[9]将博弈论用于社区发现，根据节点的相似度提出收益函数，并提出两种节点的更新策略。但是这些已有的算法均没有考虑节点自身属性在博弈中收益及节点博弈顺序的影响。

2 G_NA 算法

基于对已有社区发现算法的分析，本文提出一种基于节点属性的社区发现博弈算法——G_NA。算法将社区划分的过程看作一种节点之间的博弈。给定一个复杂网络，假设每个节点是一个自私的参与者。在社区形成过程中，每个节点的重要性不同，决定其选择策略的顺序不同，且对社区形成过程中的收益贡献值不同。因此选择一种节点属性初始化社区，节点按照一定的顺序自己选择策略提高收益。社区结构可以解释为博弈达到均衡时的状态。算法允许每个节点选择多个社区，这样可以划分重叠社区。

给定静态网络图 $G=(V,E)$，其中：$n=|V|$，$m=|E|$。假设 G 是无向无权网络。本文中，集合 V 中的元素称为节点或代理。每个节点选择其想加入的社区集合。所有可能的社区集合表示为 $[k]=\{1,2,...,k\}$。值得注意的是，最终得到的社区个数可能远小于 k。

2.1 算法的收益函数

社会网络中，节点加入某个社区使得自身利益提高的同时，会付出相应的代价。例如，员工由原来公司跳槽至现在公司，在工资提高的同时，会伴随着原来公司的违约赔偿。因此，节点的收益由增益函数和损失函数共同决定。本文基于 Newman 模块度，并加入节点度值对社区收益的贡献比例，提出增益函数。节点的收益是节点对模块度的贡献量。因此，基于 Newman 模块度提出增益函数，在增益函数中加入节点度值占网络中节点度值总和的比例，可以更好地描述节点的收益。

基于 Newman 模块度，并加入节点度值的比例。本文提出的第 i 个节点的增益函数为

$$g_i(L)=\frac{1}{2m}\sum_{j\neq i,j\in[n]}\left(A_{ij}\hat{\delta}(i,j)-p_{ij}\frac{d_i d_j}{2m}\cdot|L_i\cap L_j|\right) \tag{4}$$

其中：当$|L_i\cap L_j|\geq 1$时，$\hat{\delta}(i,j)$=1；否则$\hat{\delta}(i,j)$=0。A 是 G 的邻接矩阵，p_{ij} 是节点 i 加入节点 j 所在社区的度值比例，p_{ij} 计算方式如下：

$$p_{ij}=\frac{d_i}{\sum_{j'\in|C(j)|}d_{j'}} \tag{5}$$

其中：$C(j)$ 表示节点 j 所在社区的节点集合；j' 表示节点 j 所在社区中的节点，包括新加入的节点 i。

本文的损失函数为 $l_i(L)=(|L_i|-1)\cdot\frac{1}{2m}$，其中：$m$ 为网络中的边数；$|L_i|$ 为节点 i 的社区标签。

因此，收益函数 $u_i(\cdot)=g_i(\cdot)-l_i(\cdot)$，即

$$u_i(L)=g_i(L)-l_i(L)=\frac{1}{2m}\sum_{j\neq i,j\in[n]}\left(A_{ij}\hat{\delta}(i,j)-p_{ij}\frac{d_i d_j}{2m}\cdot|L_i\cap L_j|\right)-\frac{|L_i|-1}{2m} \tag{6}$$

前面已经介绍过，潜在博弈是允许纯纳什均衡的一类博弈。在潜在博弈中，在节点的策略选择中定义了相关的潜在函数 $\Phi(\cdot)$。如果对每个策略空间 L 和每个节点 v_i 的策略 L_i' 有 $\Phi(L)-\Phi(L_{-i},L_i')=u_i(L_{-i},L_i')-u_i(L)$，则社区形成博弈是潜在博弈。显然，本文提出的收益函数关于 1/2 线性，所以博弈为潜在博弈，可以达到纳什均衡。

2.2 节点的策略

在本文的社区形成博弈中，节点 $v_i\in V$ 的策略是其想加入的社区的子集，也就是 $[k]$ 的子集。定义 $L_i\subseteq[k]$ 为节点 $v_i\in V$ 的策略，也称之为节点 $v_i\in V$ 的社区标签。允许 $L_i=\varnothing$，也就是节点 $v_i\in V$ 可以不属于任何社区。定义 $L=(L_1,L_2,\ldots,L_n)$ 为策略空间，也就是所有节点的社区标签的一个向量。

令除 i 之外节点的社区标签为 L_{-i}，用 (L_{-i},L_i') 表示策略空间，其中 L 的第 i 个元素由 L_i' 代替。定义 $v_i\in V$ 的收益函数为 $u_i(L)=g_i(L)-l_i(L)$。在社区形成博弈中，给定其余节点的策略 L_{-i}，节点 $v_i\in V$ 的最优策略为 $\arg\max_{L_i'\subseteq[k]} g_i(L_{-i},L_i')$。

给定任一网络，计算每个节点的重要度，按照重要度值从大到小排列节点，令节点选择策略，直到收益值不能变大。节点的策略有：

a）加入：节点通过在 L_i 中加入一个新的标签表示它加入除它已经所在社区之外的社区。

b）离开：节点通过在 L_i 中移除一个标签表示它离开该社区。

c）转换：节点通过在 L_i 中替换一个标签表明它从一个社区转换至另一个社区。

节点的策略分别如图 1~3 所示。每个图中虚线分隔的节点组成不同的社区，从左到右编号为 1，2。

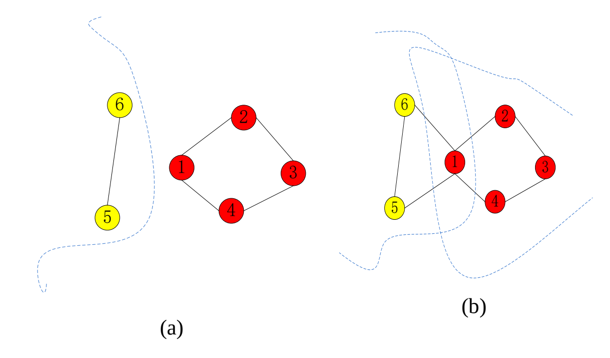


图 1 节点加入社区策略示意图

Fig.1 Schematic diagram of node joining community

图 1 表示节点的加入社区策略。在图 1（a）中节点 1 属于社区 2，其标签为 L_1={2}；图 1（b）中，当节点 1 加入社区 1 后，其标签为 L_1={2,1}。

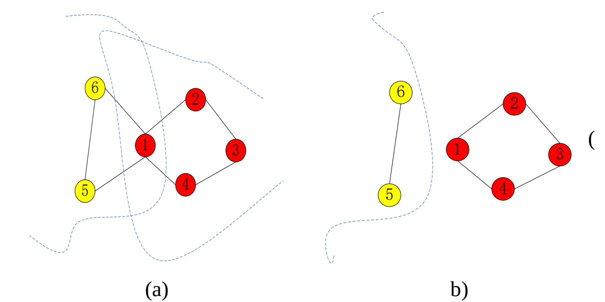


图 2 节点离开社区策略示意图

Fig.2 Schematic diagram of node leaving community

图 2 表示节点的离开社区策略。在图 2（a）中节点 1 属于社区 1，2，其标签为 L_1={1，2}；图 2（b）中，当节点 1 离开社区 1 后，其标签为 L_1={2}。

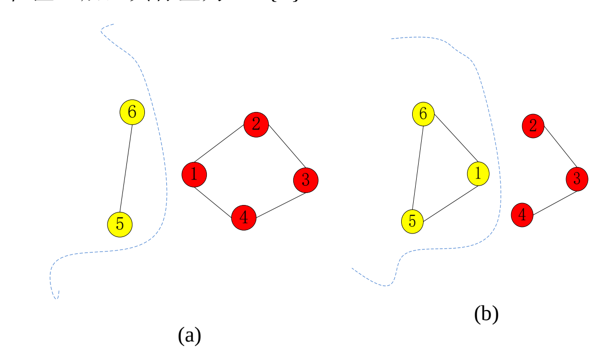


图 3 节点转换社区策略示意图

Fig.3 Schematic diagram of node converting in communities

图 3 表示节点的转换社区策略。在图 3（a）中节点 1 属于社区 2，其标签为 L_1={2}；图 3（b）中，当节点 1 由社区 2 转换到社区 1 后，其标签为 L_1={1}。

2.3　算法描述

本文首先对节点进行重要度排序，进而依次选择策略更新收益。节点重要度用于度量节点在整个网络中的影响力，本文采用 Zhang 等人[10]提出的基于贝叶斯网络的用户节点重要度计算方法归一化网络中所有节点求出的重要度。该重要度是基于先验属性得到的重要度，仅仅根据节点先验属性计算得到的归一化影响力是不够的，因为网络中节点和节点之间存在紧密的联系，一个拥有更多重要度节点链接的节点会更加重要，节点重要度越大，对其他节点的影响力越大，节点越有权利先进行策略选择。重要度计算公式如下：

$$NI(i) = Inf(i) + \alpha * \sum_{j \in N(i)} \frac{Inf(j)}{d(j)} \tag{7}$$

其中：$NI(i)$ 表示节点 i 的重要度；$Inf(i)$ 表示节点 i 的先验重要度；α 表示衡量邻接节点重要度对节点 i 影响力的系数，取值为 0~1，在 Zhang 等人的论文《Label propagation algorithm for community detection based on node importance and label influence》中 α =0.4 时，实验效果最佳，因此本文实验过程中 α 取值为 0.4，此时式（7）为 $NI(i) = Inf(i) + 0.4 * \sum_{j \in N(i)} \frac{Inf(j)}{d(j)}$ 。其中：$N(i)$ 表示节点 i 的邻接节点集；$d(j)$ 表示节点 i 的邻接节点 j 的度数。

本文提出基于节点属性的社区发现博弈算法——G_NA。算法设计很简单，首先初始化每个节点为单独社区，计算节点重要度，并按重要度值从大到小排序；然后设置算法迭代次数为 1，当算法迭代次数小于 2 000 时，节点依次选择策略更新收益；最后输出网络划分结果及模块度值。算法主要步骤如算法 1 所示。

算法 1　G_NA 算法步骤

算法：基于节点属性的社区发现博弈算法（G_NA）

输入：网络 G=(V, E)。

输出：网络社区结构，以及对应的模块度值。

a) begin

b) 初始化每个节点为单独社区，初始化收益 $u_i = 0$；

c) 计算每个节点的重要度值 *NI(i)*；

d) 将节点按照重要度值从大到小排序，得到节点序列；

e) do

f) 迭代次数 *t=1*；

g) 节点选择策略并根据式（6）计算收益；

h) 选择使节点收益最大的策略，更新节点的社区标签；

i) while 直到算法迭代至 2000 次；

j) 得到社区划分结果和模块度值；

k) end

3　实验分析

研究者通常在真实网络和仿真网络上验证社区发现算法的效果。为了验证本文提出算法的性能，本文分别在真实世界网络和人工合成网络上进行实验。

3.1　真实网络上的社区发现实验

3.1.1 实验数据集

本文在 Zachary 的空手道俱乐部网络（Karate[11]）、海豚网络（Dolphins[12]）、学校网络（Friendship[13]）、美国政治书籍网络（Polbooks[14]）和美国大学生橄榄球联盟网络（Football[15]）共五个网络上进行实验。每个网络的节点数和边数如表 1 所示。

表 1　真实网络节点数和边数

Table 1 Number of nodes and edges of real-world networks

网络	节点数	边数
Karate	34	78
Dolphins	62	159
Friendship	69	220
Polbooks	105	882
Football	115	613

表 1 是本文所用的真实网络的节点数和边数。从上到下，网络节点数增多，边数增多，网络规模变大。

3.1.2 实验结果评价指标

本文采用 Newman[14]提出的模块度 Q 评价社区发现效果。2006 年，Newman 提出用模块度 Q 评价未知社区结构的网络的社区划分结果。对于每一种社区划分结果，都有一个模块度 Q 值来判断划分结果的好坏。它的物理含义是社区内节点之间连接的边数与随机情况下两个节点之间连接边数的差。对于无权图可以理解为社区内部边的度数和减去社区内节点的总度数。表达式为

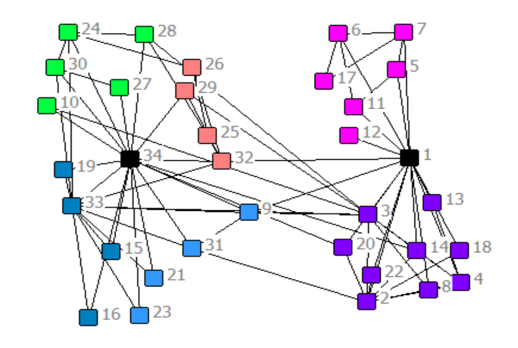
$$Q = \frac{1}{2m}\sum_{i,j}\left(A_{ij} - \frac{d_i d_j}{2m}\right)\delta\left(c_i, c_j\right) \tag{8}$$

其中：A_{ij} 为邻接矩阵，表示节点 i 与 j 之间边的权值，对于无权图，当网络中两节点之间有连接时，A_{ij} =1，否则 A_{ij} =0； d_i 表示节点 i 的度；c_i 表示 i 所属的社区；m 表示网络中的边数；$\delta\left(c_i, c_j\right)$ 是示性函数，当节点 i、j 属于同一社区时，$\delta\left(c_i, c_j\right)$ =1，否则 $\delta\left(c_i, c_j\right)$ =0。节点 j 连接到任意节点的概率是 $\frac{d_j}{2m}$，因此随机情况下，节点 i 与 j 之间的边为 $\frac{d_i d_j}{2m}$ 。

Q 的取值是[-1/2，　1]。当 Q 越接近 1，说明社区划分越明显。Q 值不仅可以衡量社区划分的准确性，也可以判断算法对于社区划分的效果。

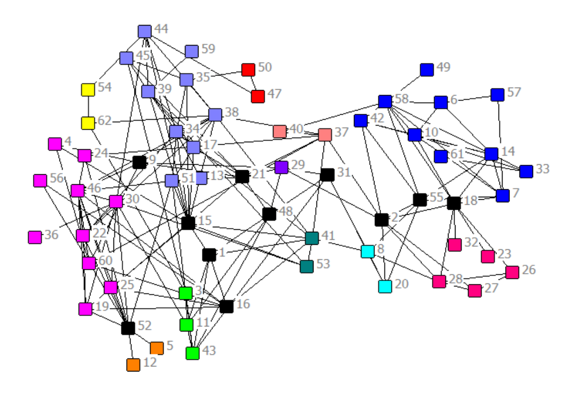
3.1.3 实验结果与分析

本文算法 G_NA 对 Karate、Dolphins、Friendship、Polbooks 以及 Football 网络的划分结果分别如图 4(a)~(e)所示。



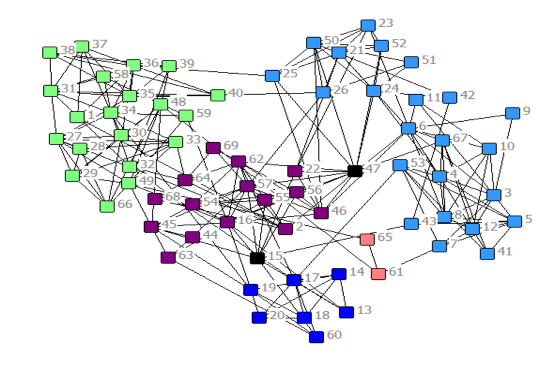
(a) G_NA 算法对 Karate 网络的社区划分示意图

(a) Communities discovered for Zachary's karate club by G_NA



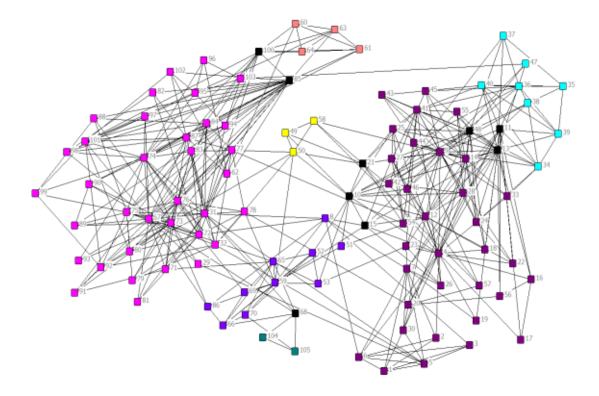
(b) G_NA 算法对 Dolphins 网络的社区划分示意图

(b) Communities discovered for Dolphins network by G_NA



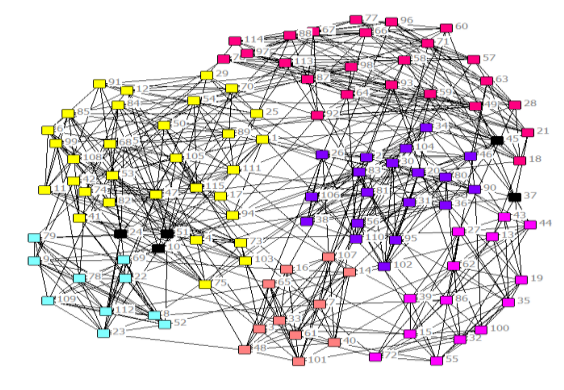
(c) G_NA 算法对 Friendship 网络的社区划分示意图

(c) Communities discovered for Friendship network by G_NA



(d) G_NA 算法对 Polbooks 网络的社区划分示意图

(d) Communities discovered for Polbooks network by G_NA



(e) G_NA 算法对 Football 网络的社区划分示意图

(e) Communities discovered for Football network by G_NA

图 4　G_NA 算法对真实网络的社区划分示意图

Fig. 4 The communities discovered for the real-world networks by G_NA

图 4(a)~(e)中分别展示了 G_NA 算法对不同真实网络的社区划分结果。如图 4（a）所示，G_NA 算法将 Karate 网络划分成 5 个社区，共发现 2 个重叠节点，分别为 1 和 34。如图 4（b）所示，G_NA 算法将 Dolphins 网络划分为 12 个社区，共发现 11 个重叠节点，分别为 1、2、9、15、16、18、21、31、48、52、55。如图 4 中 c）所示，G_NA 算法将 Friendship 网络划分为 5 个社区，共发现 2 个重叠节点，分别为 15 和 47。如图 4（d）所示，G_NA 算法将 Polbooks 网络划分为 7 个社区，共发现 9 个重叠节点，分别为 10、11、13、15、21、48、68、85、100。如图 4（e）所示，G_NA 算法将 Football 网络划分为 6 个社区，共发现 5 个重叠节点，分别为 10、24、37、45、51。（重叠节点为图中黑色节点所示）

可以看出 G_NA 算法与以前的研究相比发现了更丰富的重叠结构。G_NA 算法可以发现网络中重叠的社区，因此可以提供关于社区结构的更多有意义的信息。从实验结果可以得出结论：G_NA 算法能够发现比以前的社区发现方法更精细的社区。这些更加细密的社区揭示了网络中更详细的社区结构，并可用于社区互联和更大社区结构的进一步研究。

本文将 G_NA 算法与 Game、LPA_NI 和 LFM[16]算法进行对比，实验结果如图 5 所示。

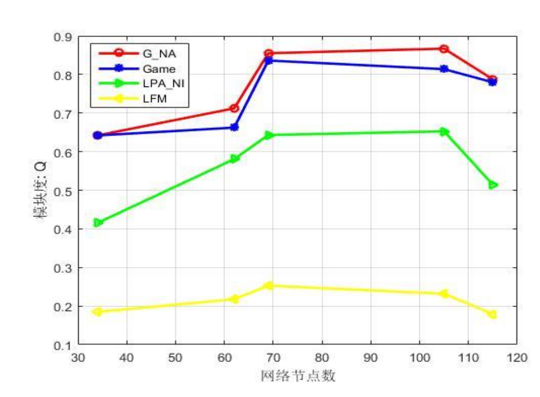


图 5　对真实网络划分的实验结果

Fig.5　Experimental results for real-world networks partitioning

由图 5 可以看出，本文算法对真实网络的划分效果优于其他算法。且基于博弈论的 Game 算法对网络的划分效果优于

LPA_NI 算法和 LFM 算法，因此将博弈论用于社区发现具有良好的研究前景和研究意义。

3.2 人工网络上的社区发现实验

3.2.1 实验数据集

本文对人工网络的社区划分实验采用的仿真网络为 LFR 基准网络。LFR 基准网络由 Lancichinetti 等人[17]提出。网络主要包括以下参数：N 表示网络中的节点数；k 表示网络中节点的平均度，maxk 表示节点的最高度数；minc 表示节点数最少的社区中的节点数，maxc 表示节点数最多的社区中的节点数；on 表示网络中所属社区大于 1 的重叠节点的个数，om 表示重叠节点所属的社区个数；mu 为混合参数，表示社区内部节点与社区外部连接的概率，随着 mu 值增大，社区发现的难度变高。LFR 基准网络的生成参数及意义如表 2 所示。可通过调节参数产生具有重叠性和层次性的网络，因此 LFR-benchmark 非常适用于重叠社区发现算法发现结果的测试与对比。本文生成非重叠 LFR 基准网络的参数设置如表 3 所示。生成重叠 LFR 基准网络的参数设置如表 4 所示。

表 2 LFR 基准网络的生成参数

Table 2 Parameters of LFR networks

参数	含义
N	网络节点数
k	节点平均度
maxk	节点最大度
minc	最小社区规模
maxc	最大社区规模
mu	混合参数
on	重叠节点数
om	重叠节点所属社区数

本文分别生成具有 200 节点、500 节点和 1 000 节点的 3 组非重叠 LFR 网络。每组 LFR 网络中混合参数分别为 0.1、0.2、0.3，如表 3 所示。

表 3 非重叠 LFR 网络的生成参数

Table 3 Parameters of non-overlapping LFR networks

网络编号	N	k	maxk	minc	maxc	mu
NO_LFR1，2，3	200	5	8	10	30	0.1，0.2，0.3
NO_LFR4，5，6	500	15	40	15	40	0.1，0.2，0.3
NO_LFR7，8，9	1000	30	50	20	50	0.1，0.2，0.3

本文分别生成重叠节点比例为 2%、4%、6%和 8%的 2 组重叠 LFR 网络。2 组网络中节点数分别为 200、1000。节点数为 200 的网络 mu 值为 0.1，节点数为 1 000 的网络 mu 值为 0.2，如表 4 所示。

表 4 重叠 LFR 网络的生成参数

Table 4 Parameters of overlapping LFR networks

网络	N	k	maxk	minc	maxc	mu	on	om
O_LFR1,2,3,4	200	5	8	10	30	0.1	4,8,12,16	2
O_LFR5,6,7,8	1000	30	50	20	50	0.2	20,40,60,80	2

3.2.2 实验结果评价指标

本文采用互信息 NMI 评价算法对人工网络的划分结果与人工网络自身社区的差异度。互信息量[17]定义为

$$NMI=\frac{I\left(\pi^{a},\pi^{b}\right)}{\sqrt{H\left(\pi^{a}\right)H\left(\pi^{b}\right)}} \tag{9}$$

$$H\left(\pi^{a}\right)=\sum_{h}^{k^{(a)}}\frac{n_{h}^{a}}{n}\log(\frac{n_{h}^{a}}{n}) \tag{10}$$

$$H\left(\pi^{b}\right)=\sum_{l}^{k^{(b)}}\frac{n_{l}^{b}}{n}\log(\frac{n_{l}^{b}}{n}) \tag{11}$$

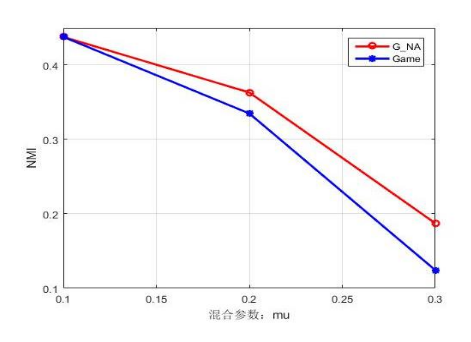
$$I\left(\pi^{a},\pi^{b}\right)=\sum_{h}\sum_{l}\frac{n_{h,l}}{n}\log(\frac{n_{h,l}}{n}\Big/(\frac{n_{h}^{a}}{n}\frac{n_{l}^{b}}{n})) \tag{12}$$

其中：π^a 、π^b 分别表示某个社区结构；$k^{(a)}$ 表示社区结构 π^a 中社区的个数；n_h^a 表示在社区结构 π^a 中第 h 个社区的节点数；$n_{h,l}$ 表示同时在 π^a 中的第 h 个社区，在 π^b 中的第 l 个社区，这样的节点的个数。NMI 的取值为[0, 1]，值越大，表示算法发现的社区结构与网络本身的社区结构一致性越高。

3.2.3 实验结果与分析

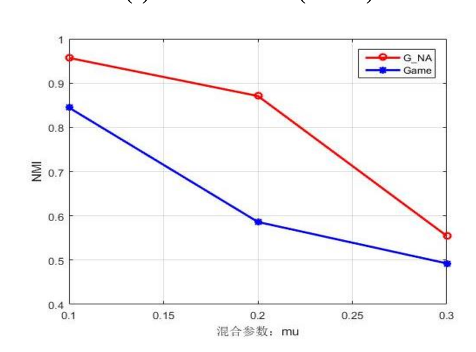
本文比较 G_NA 算法与 Game 算法对人工网络的划分结果。算法对非重叠网络的划分结果如图 6 所示。

图 6（a）~（c）分别为算法在具有 200 个节点、500 个节点和 1 000 个节点的网络上的划分结果。由图 6（a）可以看出，当 mu=0.1 时，G_NA 算法的划分结果与 Game 算法的划分结果接近；由（c）可以看出，当 mu=0.1 时，Game 算法的划分结果优于 G_NA 算法的划分结果，但是从（a）~（c）可以看出，当 mu 值增大时，G_NA 算法的划分结果明显优于 Game 算法。因此可以说明 G_NA 算法对于社区结构不明显的网络划分有很大优势。



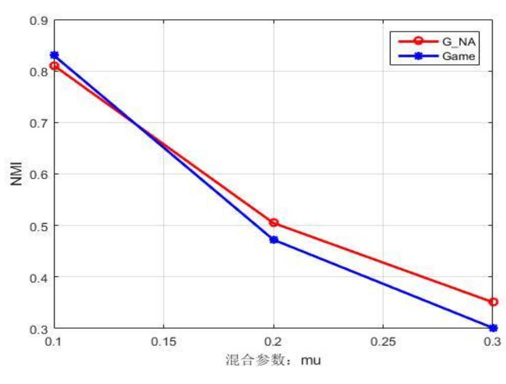
（a）网络划分结果（N=200）

(a) Detection result (N=200)



（b）网络划分结果（N=500）

(b) Detection result (N=500)



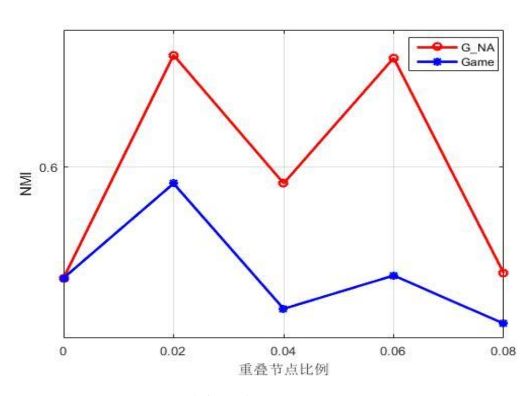
（c）网络划分结果（N=1000）

(c) Detection result (N=1000)

图 6　算法对非重叠网络的划分结果

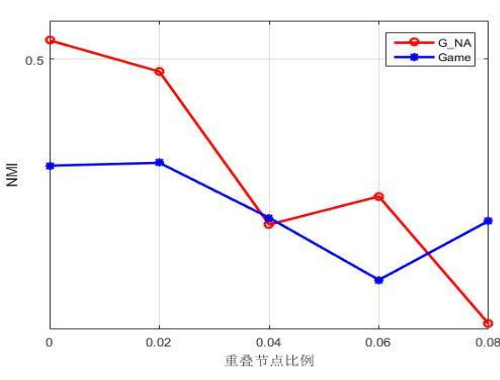
Fig.6 Detection results of non-overlapping networks

算法对重叠网络的社区划分结果如图 7 所示。



(a)网络划分结果(N=200,mu=0.1)

(a) Detection result (N=200, mu=0.1)



(b)网络划分结果(N=1000,mu=0.2)

(b) Detection result (N=1000, mu=0.2)

图 7　算法对重叠网络的划分结果

Fig.7　Detection results of Overlapping networks

图 7（a）（b）分别为算法在 N=200，mu=0.1 和 N=1000，mu=0.2 的网络上的划分结果。图中横轴表示重叠节点占总节点的比例。由图 7（a）可以看出，G_NA 算法的划分结果优于 Game 算法的划分结果；在图（b）中，当重叠节点数为 0，20、40 时，G_NA 算法的划分结果优于 Game 算法的划分结果；当重叠节点数为 60 时，G_NA 算法的划分结果略逊色于 Game 算法的划分结果；当重叠节点数为 80 时，G_NA 算法的划分结果与 Game 算法的划分结果差距较大。因此，总的来说，G_NA 算法的划分结果优于 Game 算法的划分结果。

4　结束语

本文在分析现有社区发现算法不足的基础上，提出基于节点属性的社区发现博弈算法。针对已有算法未考虑节点属性对节点策略选择影响的局限性，提出具有节点度值比例的增益函数；针对已有算法未考虑节点属性对节点在网络中选择顺序影响的不足，提出基于节点重要度排序的博弈论社区发现算法；在基于节点重要度初始化的博弈论社区发现算法中，节点按照重要度从大到小排序，并依次选择加入社区、离开社区或转换社区的策略提高收益，直到所有节点不能增大其收益。

在接下来的工作中将更多地结合博弈论与社区发现的本质提出更适合社区发现的收益函数。另外在对本文算法的实验过程中发现节点更多地考虑加入自身邻接节点所在的社区，因此在接下来的工作中可以设置节点的策略为其邻接节点社区的选择。

参考文献：

[1] Kunegis J. Social network datasets [M]. New York: Springer, 2017.

[2] 姚莹. 基于遗传优化的复杂网络社区检测技术研究 [D]. 南京. 南京邮电大学，2017. (Yao Ying. Research on complex network community detection technology based on genetic optimization [D]. Nanjing: Nanjing University of Posts and Telecommunications, 2017.)

[3] 郭雷，许晓鸣. 复杂网络 [M]. 上海：上海科技教育出版社, 2006. (Guo Lei, Xu Xiaoming. Complex network [M]. Shanghai: Shanghai Science and Technology Education Press, 2006.)

[4] Wei Chen, Zhenming Liu, Xiaorui Sun, *et al.* A game-theoretic framework to identify overlapping communities in social networks [J]. Data Mining & Knowledge Discovery, 2010, 21 (2): 224-240.

[5] Rubinstein, Ariel. A course in game theory [M]. [S. l.] : MIT Press, 1994.

[6] Aumann R J, Brandenburger A. Epistemic conditions for Nash equilibrium [M]// Readings in Formal Epistemology. [S. l.] : Springer International Publishing, 2016: 113-136.

[7] Monderer D, Shapley L S. Potential games [J]. Games & Economic Behavior, 1996, 14 (1): 124-143.

[8] Alvari H, Hashemi S, Hamzeh A. Detecting overlapping communities in social networks by game theory and structural equivalence concept [M]. Berlin: Springer, 2011.

[9] Zhou Xu, Zhao Xiaohui, Liu Yanheng, *et al.* A game theoretic algorithm to detect overlapping community structure in networks [J]. Physics Letters A, 2018, 382 (13): 872-879.

[10] Zhang Xiankun, Ren Jing, Song Chen, *et al.* Label propagation algorithm for community detection based on node importance and label influence [J]. Physics Letters A, 2017, 381 (33): 2691-2698.



[11] Zachary W W. An information flow model for conflict and fission in small groups [J]. Journal of Anthropological Research, 1977, 33 (4): 452–473.

[12] Lusseau D, Schneider K, Boisseau O J, *et al.* The bottlenose dolphin community of doubtful sound features a large proportion of long-lasting associations [J]. Behavioral Ecology and Sociobiology, 2003, 54 (4): 396-405.

[13] Yang Liang, Cao Xiaochun, He Dongxiao, *et al.* Modularity based community detection with deep learning [C]// Proc of International Joint Conference on Artificial Intelligence. [S. l.] : AAAI Press, 2016: 2252-2258.

[14] Newman M E J. Modularity and community structure in networks [J]. Proceedings of the National Academy of Sciences, 2006, 103 (23): 8577-8582.

[15] Girvan M, Newman M E J. Community structure in social and biological networks [J]. Proc. Natl Acad. Sci. USA. 2002, 99 (12): 7821-7826.

[16] Lancichinetti A, Fortunato S, Kertész J. Detecting the overlapping and hierarchical community structure of complex networks [J]. New Journal of Physics, 2008, 11 (3): 19-44.

[17] Lancichinetti A, Fortunato S. Benchmarks for testing community detection algorithms on directed and weighted graphs with overlapping communities [J]. Physical Review E, 2009, 80 (1): 16118.